Lec :5

Multiple Features

Linear regression with multiple Variables

Size (Feet^2)	no.of bedrooms	no.of Floors	Age of home (Year)	Price ($1000)
X_1	X_2	X_3	X_4	Y
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
:	:	:	:	:

n: no. of Features.

$x^{(i)}$: input (Features.) of i^{th} training example.

$x_j^{(i)}$: Value of Feature j in i^{th} training example.

ex: From the table $n = 4$

$x^{(i)} = X_1 , X_2 , X_3 , X_4$

$x_j^{(i)} = 2104 , 3 , 1 , 45$

Hypothesis :

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n.$$

For convenience of notation, define $x_0 = 1$

Parameters : $\theta_0, \theta_1, \ldots \theta_n$

Cost Function: $J(\theta_0, \theta_1, \ldots \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$

Gradient descent : $(n \geq 1)$

Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \dots, \theta_n)$$

}

(Simultaneously update for every $j = 0, \dots, n$)

$$\rightarrow \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

$$\rightarrow \theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_1^{(i)}$$

$$\rightarrow \theta_2 := \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_2^{(i)}$$

...

Feature Scaling

Idea: Make sure features are on a similar scale.

[1] Scale by dividing the actual feature value by the feature maximum value.

ex: $x_1 = \text{size } (0 - 2000 \text{ feet}^2) \rightarrow x_1 = \frac{\text{Size (feet}^2)}{2000}$

$x_2 = \text{number of bedrooms } (1-5) \rightarrow x_2 = \frac{\text{no. of bedrooms}}{5}$

get every feature into approximately

$$0 \leq x_i \leq 1$$

range.

القيمة الاصلية · Average · مجموعهم / عددهم

$\boxed{2}$ Scale by replacing x_i by $\dfrac{x_i - \mu_i}{S}$ where $S = max - min$.

Max - Min

S: Standard deviation.

μ_i: average mean.

→ replace x_i with $x_i - \mu_i$ to make features have approximately Zero mean (Do not apply to $x_0 = 1$).

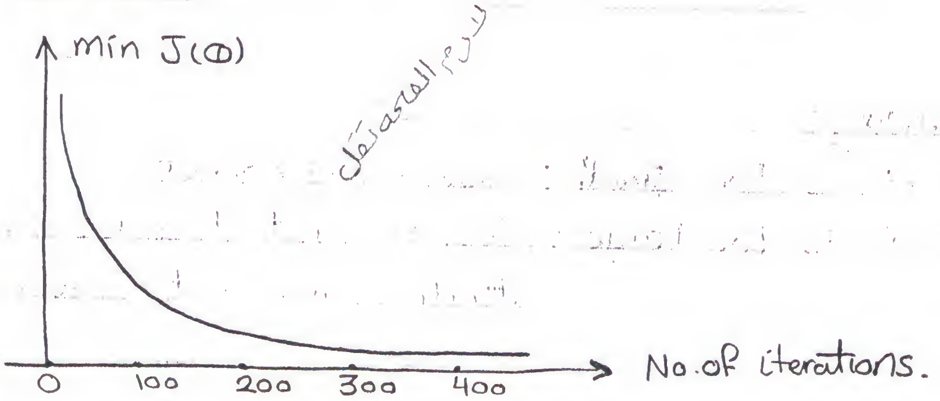
ex: $x_1 = \dfrac{\text{Size} - 1000}{2000} \rightarrow -0.5 \leqslant x_1 \leqslant 0.5$

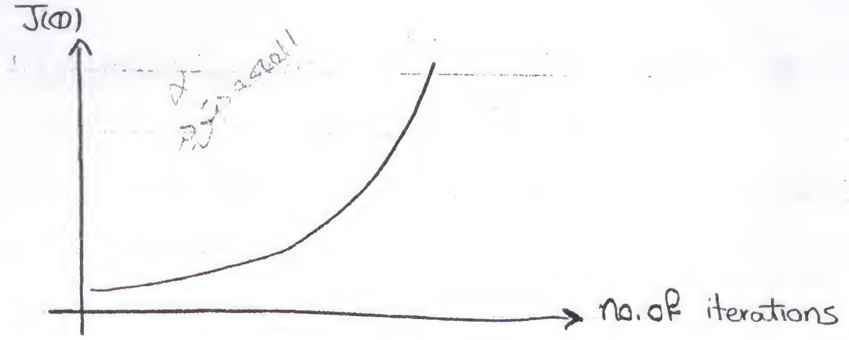
$x_2 = \dfrac{\#\text{bedroom} - 2}{5} \rightarrow -0.5 \leqslant x_2 \leqslant 0.5$

———

<u>Making Sure gradient descent is working Correctly.</u>

1. Plot $J(\Theta)$ as a function of the number of iterations.
2. It works Correctly when $J(\Theta)$ <u>decreases</u> after every iteration.

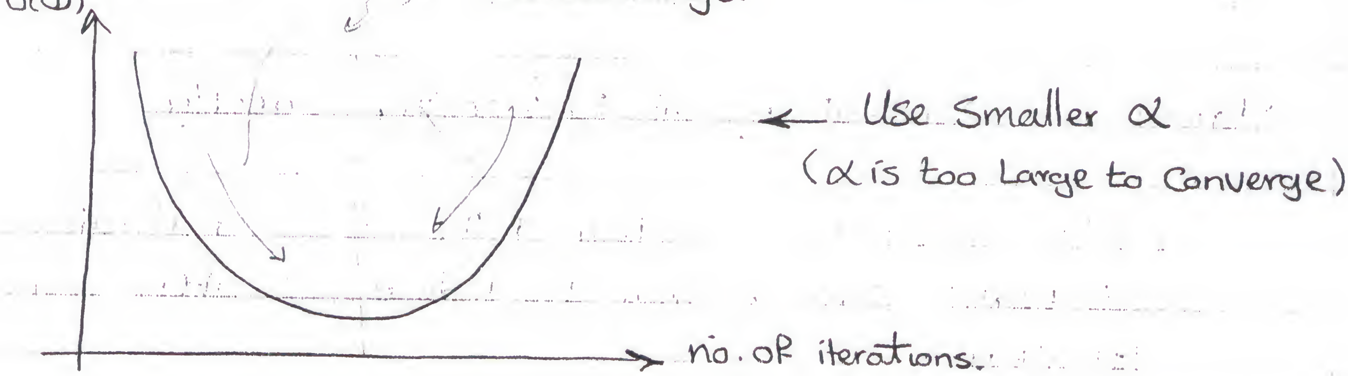




Here, Gradient descent not working.

3. For sufficienty small α, $J(\theta)$ should decrease on every iteration.

. But if α is too small, gradient descent can be slow to converge.



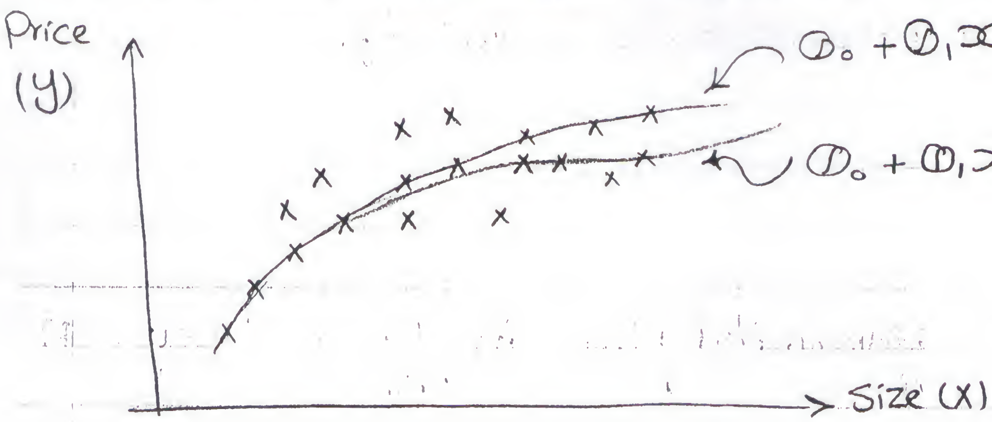
Summary:

- If α is too small : slow convergence.
- If α is too Large : $J(\theta)$ may not decrease on every iteration; may not converge.

To choose α, Try

...., 0.001, ..., 0.01, ..., 0.1, ... 1, ...

Features and <u>Polynomial</u> Regression.



$\theta_0 + \theta_1 x_1 + \theta_2 x_2^2$

$\theta_0 + \theta_1 x_1 + \theta_2 x_2^2 + \theta_3 x_3^3$.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$h_\theta(x) = \theta_0 + \theta_1 (\text{size}) + \theta_2 (\text{size})^2 + \theta_3 (\text{size})^3$$ لو Feature واحد (Size)

$x_1 = \text{Size}$, $x_2 = (\text{Size})^2$, $x_3 = (\text{size})^3$.

- <u>Normal Equation</u>

  : method to solve for θ analytically.

x_0	Size (feet^2) x_1	No.of bedrooms x_2	No.of floors x_3	Age of home x_4	Prise (1000\$) y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178
1	3000	4	1	38	5

لتطبيق امر المعطى الرئيسي

$$\theta = (X^T X)^{-1} X^T y$$

الصفوف عدد الأعمدة

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \\ 1 & 300 & 4 & 1 & 38 \end{bmatrix} , \quad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \\ 540 \end{bmatrix}$$

$\longrightarrow$ m training examples (عدد الصفوف)

$\longrightarrow$ n features (عدد الأعمدة)

Gradient Descent	Normal Equation.
• Need to Choose α	• No need to Choose α.
• Needs many iterations.	• Don't need to iterate.
• <u>Works well even when n is Large.</u>	• Need to Compute $(X^T X)^{-1}$
	• <u>Slow if n is very Large.</u> لو m (training set) كبيرة